DNA to be detected; recovered faecal DNA can be amplified more than a billion-fold by polymerase chain reaction before it is measured.

Early investigations targeting single mutations, usually K-ras, show that mutations in tumours can be detected in stools from the same patients.[8–10] Since colorectal neoplasms are genetically heterogeneous, however, no one mutation has been identified that is universally expressed. Mutant K-ras, for example, is present in less than half of all colorectal neoplasms; this would restrict the maximum sensitivity of this test for colorectal cancer to less than 50% if it was used as the sole marker for screening in a stool assay.[8–10] Also, since mutant K-ras may arise from non-neoplastic sources, such as pancreatic hyperplasia, this marker may lack specificity. Multiple DNA alterations should be targeted to achieve high rates of detection, and each component marker must be specific for a neoplasm to avoid false positive results.

Data from pilot projects suggest that the diagnostic yield improves when a stool assay with multiple targets is directed at a spectrum of DNA alterations commonly expressed by cancers.[11] The assay used in the pilot study included 15 mutational "hot spots" on K-ras, p53, and APC genes; BAT-26, a microsatellite instability marker (a genomic alteration present in 15-20% of colon cancers), and long (non-apoptotic) DNA. Using one stool per patient, which was tested blind, DNA alterations were detected in 20 of 22 (91%) patients with colorectal cancer, 9 of 11 (82%) patients with adenomas >1 cm, and 2 of 28 (7%) controls who had had normal colonoscopies. When K-ras was dropped from the assay, sensitivity for cancer was unaffected but it fell to 73% for large adenomas and specificity rose to 100%. Larger studies are clearly indicated to corroborate these early outcomes.

Preliminary data suggest that components of this assay panel may also detect cancers that occur above the colon, including in the lung, at sensitivities comparable to those for colorectal neoplasia.[12]

Thus, stool screening with DNA markers could have benefits beyond detecting colorectal neoplasms: it may be useful in controlling other cancers as well. It will be important to evaluate the implications of these findings using screening algorithms and evaluations of overall cost effectiveness.

Assay methods are labour intensive and must be streamlined for large scale testing, but exciting technologies are emerging to make this possible. While it is too early to know for certain, molecular markers may improve the effectiveness and efficiency of stool screening.

David A Ahlquist *professor of medicine*

Division of Gastroenterology and Hepatology, Mayo Clinic E-19, Rochester, MN 55905, USA (ahlquist.david@mayo.edu)

DAA has been given partial research support (to help with supplies and assay costs) through a grant from EXACT Laboratories, which developed the multitarget assay system discussed in this editorial.

1 Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota colon cancer study. *N Engl J Med* 1993;328: 1365-71.
2 Hardcastle JD, Chamberlain JO, Robinson MN, Moss SM, Amar SS, Balfour TW, et al. Ramdomised controlled trial of faecal-occult blood screening for colorectal cancer. *Lancet* 1996;348:1472-7.
3 Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomized study of screening for colorectal cancer with fecal-occult blood test. *Lancet* 1996;348:1467-71.
4 Ahlquist DA. Fecal occult blood testing for colorectal cancer: can we afford to do this? *Gastroenterol Clin N Am* 1997;26:41-55.
5 Ahlquist DA, Harrington JJ, Burgart LJ, Roche PC. Morphometric analysis of the "mucocellular layer" overlying colorectal cancer and normal mucosa: relevance to exfoliated stool screening markers. *Hum Pathol* 2000;31:51-7.
6 Loktionov A, O'Neill IK, Silvester KR, Cummings JH, Middleton SJ, Miller R. Quantitation of DNA from exfoliated colonocytes isolated from human stool surface as a novel noninvasive screening test for colorectal cancer. *Clin Cancer Res* 1998;4:337-42.
7 Boland CR, Sato J, Saito K, Carethers JM, Marra G, Laghi L, et al. Genetic instability and chromosomal aberrations in colorectal cancer: A review of current models. *Cancer Detect Prev* 1998;22:377-82.
8 Sidransky D, Tokino T, Hamilton SR, Kinzler KW, Levin B, Frost P, et al. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* 1992;256:102-5.
9 Eguchi S, Kohara N, Komuta K, Kanematsu T. Mutations of the p53 gene in stool of patients with resectable colorectal cancer. *Cancer* 1996;77: 1707-10.
10 Villa E, Dugani A, Rebecchi AM, Vignoli A, Grottola A, Buttafoco P, et al. Identification of subjects at risk for colorectal carcinoma through a test based on K-ras determination in the stool. *Gastroenterology* 1996;110:1346-53.
11 Ahlquist DA, Skoletsky JE, Boynton KA, Harrington JJ, Mahoney DW, Pierceall WE, et al. Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay system. *Gastroenterology* (in press).
12 Ahlquist DA, Cameron AJ, Jett J, Pearson RK, de Groen PC, Edell E, et al. Universal detection of aerodigestive cancers by assay of nonapoptotic human DNA in stool [abstract]. *Gastroenterology* 2000;188:A855.

Which clinical studies provide the best evidence?

The best RCT still trumps the best observational study

A common question in clinical consultations is: "For this person, what are the likely effects of one treatment compared with another?" The central tenet of evidence based medicine is that this task is achieved by using the best evidence combined with consideration of that person's individual needs.[1] A further question then arises: "What is the best evidence?" Two recent studies in the *New England Journal of Medicine* have caused uproar in the research community by finding no difference in estimates of treatment effects between randomised controlled trials and non-randomised trials.

The randomised controlled trial and, especially, systematic reviews of several of these trials are traditionally the gold standards for judging the benefits of treatments, mainly because it is conceptually easier to attribute any observed effect to the treatments being compared. The role of non-randomised (observational) studies in evaluating treatments is contentious: deliberate choice of the treatment for each person implies that observed outcomes may be caused by differences among people being given the two treatments, rather than the treatments alone. Unrecognised confounding factors can always interfere with attempts to correct for identified differences between groups.

These considerations have supported a hierarchy of evidence, with randomised controlled trials and derivatives at the top, controlled observational studies

BMJ 2000;321:255–6

in the middle, and uncontrolled studies and opinion at the bottom. The best evidence to use in decisions is then the evidence highest in the hierarchy. Evidence from a lower level should be used only if there is no good randomised controlled trial to answer a particular clinical question. This view was supported by two studies that found larger effects in observational studies than in randomised controlled trials of the same treatment comparisons.[2 3]

However, these findings were not confirmed by the two latest studies in the *New England Journal of Medicine*, which compared individual randomised controlled trials with observational studies in 19 therapeutic areas[4] and meta-analyses of randomised controlled trials with meta-analyses of cohort and case-control studies in five therapeutic areas.[5] No major differences were found between the estimates of treatment effects in the observational studies and randomised controlled trials.

Do these newer results overturn the idea of best evidence and mean that we should abandon the use of a hierarchy of evidence? The authors speculate that their latest comparisons of study designs failed to confirm older studies for two main reasons. Firstly, observational studies have improved (people who are given different treatments may be more comparable or researchers may be better at allowing for residual differences), and secondly, earlier comparisons used particularly poor observational designs (such as historical controls that use control data from a different set of people and from an earlier period than the one used for the treatment being studied).

However, an accompanying editorial[6] found three additional problems with the latest comparisons of observational studies and randomised controlled trials.[4 5] Firstly, the search for corresponding randomised controlled trials and observational studies in well known journals selected a small, potentially atypical subgroup of available randomised controlled trials. Conclusions based on the selected therapies might not extend to other areas. Secondly, one observational study did not involve any treatment but explored risk factors in the general population. Thirdly, meta-analyses and randomised controlled trials published after the studies in the *New England Journal of Medicine* did not follow the same pattern and disagreed with results of corresponding observational studies. For example, a new meta-analysis of breast cancer screening that included weighting by quality of randomised controlled trial found no evidence of benefit, in contrast to results from observational studies,[7] and a randomised controlled trial of hormone replacement therapy in menopausal women found no secondary prevention of coronary risk or reduced fracture risk, in contrast to numerous observational studies.[8]

Even before the papers in the *New England Journal of Medicine* an earlier systematic review also found no consistent difference between randomised controlled trials and observational studies in estimates of the effects of treatment in 22 areas.[9] Differential quality of care, selection of people with a larger capacity to benefit, and publication bias against negative results from observational studies could explain larger treatment effects in either study design.

The issue is further confused by another systematic review published in *JAMA* that compared eight randomised controlled trials with non-randomised trials of the same intervention and found larger effects in five of the non-randomised trials.[10]

It is not surprising that high quality randomised controlled trials and high quality observational studies can sometimes produce similar answers. Not all observational studies are misleading. The hierarchy of evidence is merely a convenient rule of thumb that, all other things being equal, randomised controlled trials are more able to attribute effects to causes. Randomised controlled trials that are well conducted remain the gold standard for evidence of efficacy. However, small inadequate ones do not automatically trump any conflicting observational study. Identifying the best evidence for any question requires detailed appraisal—for example, relevance, allocation concealment (ensuring that the assignment of interventions are unpredictable by all involved in the trial until the point of allocation), intention to treat analysis, and relevant outcomes. If high quality randomised controlled trials exist for a clinical question then they trump any number of observational studies. Limited randomised controlled trials need other forms of evidence to be appraised and considered.

A similar debate took place centuries ago in English law. The legal "best evidence rule" initially created a rigid hierarchy of evidence (that original written documents took precedence over oral evidence). It was replaced by the flexible principle that the weight given to each bit of evidence should be determined by a detailed appraisal of the characteristics of that evidence.[11]

The new studies do not justify a major revision of the hierarchy of evidence, but they do support a flexible approach in which randomised controlled trials and observational studies have complementary roles. High quality observational studies may extend evidence over a wider population and are likely to be dominant in the identification of harms and when randomised controlled trials would be unethical or impractical.

Stuart Barton *executive editor, Clinical Evidence*

BMA House, Tavistock Square, London WC1H 9JR
(sbarton@bmjgroup.com)

Competing interests: SB edits a journal that uses a flexible hierarchy of evidence.

1 Sackett DI, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.
2 Chalmers TC, Matta RJ, Smith H Jr, Kunzler A-M. Evidence favouring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977;297;1091-6.
3 Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982;72:233-40.
4 Benson K, Hartz AJ. A comparison of observational and randomized controlled trials. *N Engl J Med* 2000;342:1878-86.
5 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies and the hierarchy of research designs. *N Engl J Med* 2000;342:1887-92.
6 Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000;342:1907-9.
7 Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;355:129-34.
8 Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998;280:605-13.
9 Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998;2(13). www.hta.nhsweb.nhs.uk (accessed July 25 2000).
10 Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
11 Twining W. *Rethinking evidence*. Evanston, IL: Northwestern University Press, 1994.

We ask all editorial writers to sign a declaration of competing interests (bmj.com/guides/confli.shtml#aut). We print the interests only when there are some. When none are shown, the authors have ticked the "None declared" box.